

Data science спеціалізація проекту Coursera

Любов Феліксівна Панченко

Кафедра фізико-технічних систем і інформатики, ДЗ «Луганський національний університет імені Тараса Шевченка»,
вул. Гоголя, 1, м. Старобільськ, 92703, Україна
lubov.felixovna@gmail.com

Анотація. *Мета дослідження:* з'ясування суті поняття «specialization» в практиці масових онлайн курсів. *Об'єкт дослідження:* процес навчання на масових відкритих дистанційних курсах. *Предмет дослідження:* особливості «data science» спеціалізації проекту Coursera. *Завдання дослідження:* взяти участь в якості студента в декількох онлайн масових курсах з категорії «data science specialization», з'ясувати сутність і структуру цієї спеціалізації, визначити її особливості, шляхи включення в традиційний навчальний процес українських вузів. *Методи дослідження:* включене спостереження, контент-аналіз, аналіз продуктів діяльності. *Результати дослідження:* спеціалізація Data science проекту Coursera – серія з 9-ти курсів, що охоплюють концепції і засоби аналізу даних, починаючи з постановки дослідницьких питань і закінчуючи публікацією результатів. Послідовність курсів завершується виконанням спеціального проекту (Capstone Project). Курси в серії пов'язані жорсткою і м'якою залежністю. Щомісяця курси повторюються. Структуру курсу становить Сілабус, короткі відеолекції, тести, реєстр оцінювання, курсові проекти, форум. Для завдань з програмування використовуються R, RStudio, Git, GitHub. *Висновки і рекомендації:* шляхи включення такої форми навчання як спеціалізація data science в традиційний навчальний процес українських вузів: оновлення навчально-методичного забезпечення дисциплін, що викладаються, організація самостійної роботи студентів з матеріалами курсів, включення в тематику кваліфікаційних робіт, використання нових засобів і методів, що вивчаються в спеціалізації, в дослідженнях аспірантів і докторантів для аналізу отриманих даних.

Ключові слова: масовий відкритий онлайн курс; data science specialization; Coursera; R.

L. F. Panchenko. The study of Coursera's data science specialization

Abstract. *Objective:* To identify the characteristics of the specialization form of the massive open online courses. *Research object:* a learning process of the massive open online courses. *Research subject:* Data science specialization of Coursera. *Research objectives:* to participate as a student in

the several online courses in “data science specialization”, to find the structure of this specialization, to determine its characteristics. *Methods:* participant observation, content analysis. *Results:* Data science specialization of Coursera project is a series of 9 courses covering concepts and tools of data analysis, from the research questions formulation to results publication. The implementation of a special Capstone Project has completed this sequence of courses. Courses are repeated once a month during a year. The courses in the specialization are related with a hard and a soft dependences. Course structure consists of syllabus, short video lectures, tests, peer assessment, course projects, forum. The software R, RStudio, Git, GitHub are used for programming assignment. *Conclusions and recommendations:* there are next ways to aggregate this form in the traditional educational process of Ukrainian universities: developing training and methodological support of disciplines, the students work organization with course materials, including the topics in qualification works, using new data analysis tools and techniques in the post graduate and post doctoral studies.

Keywords: massive open online courses; Data science specialization; Coursera; R.

Affiliation: Department of physical-technical systems and informatics, Luhansk Taras Shevchenko National University, 1, Gogol Sq., Starobilsk, 92703, Ukraine.

E-mail: lubov.felixovna@gmail.com.

В останні роки спостерігається неослабний інтерес до феномену масових онлайн дистанційних курсів взагалі і до курсів, що присвячені різним аспектам аналізу даних, зокрема [1; 4; 5; 9; 11; 12] (рис. 1, рис. 2).



Рис. 1. Google Trends: пошуковий запит «Massive Open On-line course»

Разом з цим, очевидна тенденція, яка полягає у розширенні типів масових онлайн курсів. Так, наприклад, поряд з усім відомими моделями сМООС і хМООС (Дж.Сіменс) деякі автори виділяють і інші типи [1; 3]:

– BOOC (a big open online course);

- DOCC (distributed open collaborative course);
- LOOC (little open online course);
- MOOR (massive open online research);
- SPOC (small, private, online course);
- SMOC (synchronous massive open online course).

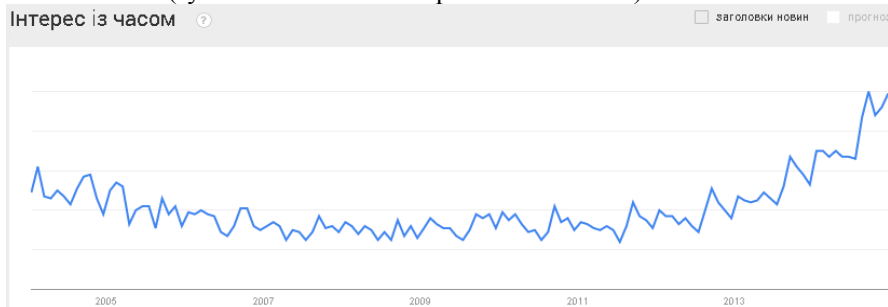


Рис. 2. Google Trends: пошуковий запит «Data Science»

Доповнимо цю таксономію ще одним видом МООС, а саме SMOOC – так звані «спеціалізації». Під «спеціалізацією», яка з'явилася порівняно нещодавно, розуміють послідовність пов'язаних між собою курсів з однієї предметної області, які надають можливість студенту глибоко вивчити предмет, і згодом застосувати свої знання на практиці у спеціальному проекті.

На кінець грудня 2014 р. у проекті Coursera було реалізовано 27 «спеціалізацій», серед яких близько п'яти присвячено галузі аналізу даних.

Серед них:

- Data Science – Johns Hopkins University;
- Data Mining – University of Illinois at Urbana-Champaign;
- Healthcare Informatics & Data Analytics – Georgia Institute of Technology;
- Become a Social Scientist: Methods and Statistics – University of Amsterdam;

Мета даного дослідження – на прикладі спеціалізації «Data science» проекту Coursera розглянути особливості цієї форми масового онлайн навчання.

Спеціалізація «Data Science» [5] проекту Cousera (рис. 3) налічує 9 курсів, які охоплюють концепції та засоби аналізу даних, починаючи з постановки питань і закінчуючи публікацією результатів. Завершується послідовність курсів виконанням спеціального проекту (Capstone Project).

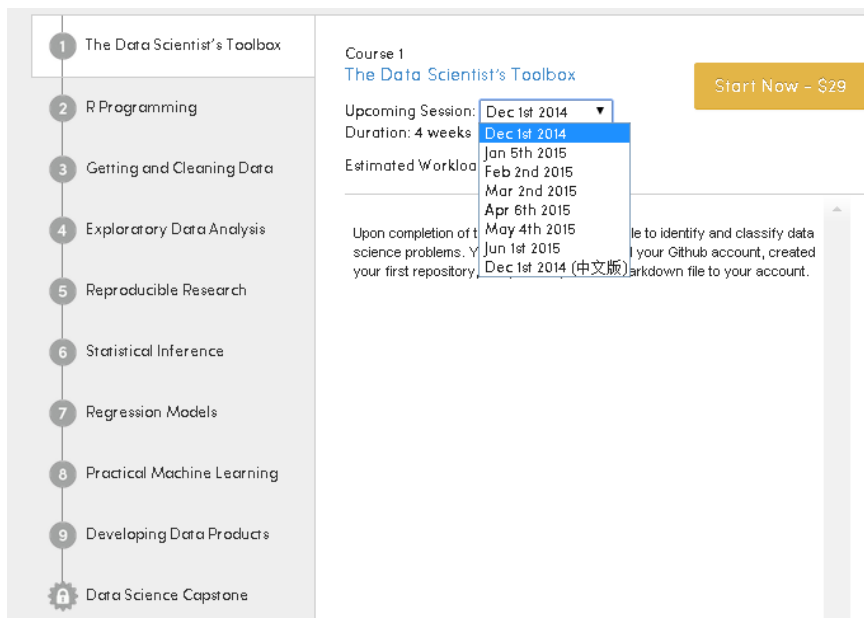


Рис. 3. «Data science» спеціалізація проекту Coursera

Дослівно, Capstone – один із завершальних або захисних каменів, які утворюють верхню частину зовнішньої кладки стіни або будівлі [2]. А стосовно сфери освіти цей термін розуміється як «вінець, точка, елемент, подія» або «останній штрих, кульмінація, вершина». Іншими словами – це глибоке дослідження в області навчальної програми, яке передбачається в заключний період навчання і ґрунтується на попередніх курсах.

Словосполучення «Capstone course» [2; 8] трактується як курс, запропонований в рамках академічної програми з метою об'єднання основних аспектів навчальної дисципліни. У багатьох американських і канадських університетах Capstone course – це курс, на якому студент навчається, щоб завершити здобуття ступеня бакалавра.

Що стосується порядку проходження курсів: творці спеціалізації виділяють два види залежності між курсами: «жорстку» і «м'яку». Жорстка залежність передбачає, що студенти в обов'язковому порядку повинні знати матеріал попередніх курсів. При м'якій залежності знання попередніх курсів вітаються і рекомендуються. Творці спеціалізації рекомендують спочатку обов'язково пройти перші два курси: «Інструменти аналітики даних», «R». Для наступних курсів описано, які курси пов'язані з іншими жорсткою або м'якою залежністю.

Для оцінювання навчальних досягнень студентів в рамках спеціалізації пропонуються традиційні інструменти Coursera: тести (quizzes), оцінювання однокурсниками один одного (peer assessment), виконання проєктів (project). Цікаво, як самі студенти оцінюють корисність цих інструментів (рис. 4).

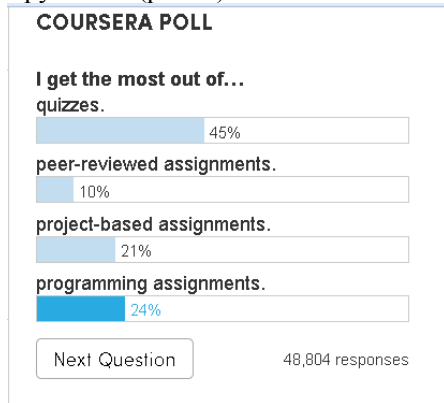


Рис. 4. Оцінювання студентами корисності елементів курсу

Можна спостерігати, що найменш корисним студенти вважають оцінювання один одного однокурсниками. Дійсно, такий вид оцінювання викликає найбільше проблем, в той же час він має потенційні соціальні можливості. Професор математики зі Стенфорду Кейт Девлін (Keith Devlin), дискутуючи про оцінку студентської роботи однокурсниками, писав: «Це як вчитися їздити на велосипеді: існує тільки один спосіб, сідайте, падаєте, сідайте, падаєте, поки щось не «кляцне» і все налагоджується. У математиці таке кляцання відбувається тільки, якщо ви студент і перевіряєте студентські роботи» [10].

Творці спеціалізації розуміють і відзначають, що рубрики для оцінювання, хоч і тестуються, але ще далекі від досконалості. Студентам пропонується спеціальний форум для обговорення питань, пов'язаних з рубриками, а також з виразом незгоди з оцінками свого проєкту однокурсниками. При цьому наголошується, що переглянути ці оцінки не є можливим, але в майбутньому це буде враховуватися для поліпшення курсів.

Особливістю спеціалізації є те, що всі проєктні завдання ґрунтуються на реальних даних, часто дуже великих.

Наведемо приклад такого проєкту з курсу «Getting and Cleaning Data». Набір даних був зібраний з метою розпізнавання виду діяльності людини за допомогою смартфона [6]. Експерименти були проведені з групою 30 добровольців у рамках вікової групи 19-48 років. Кожна

людина виконував шість видів діяльності (ходьба, ходьба вгору по сходах, спуск по сходах, сидіння, стояння і лежання) зі смартфоном Samsung Galaxy SII, закріпленим на талії. Завдяки вбудованому акселерометру і гіроскопу, експериментатори визначили 3-х осьове лінійне прискорення і 3-х осьову кутову швидкість. Сигнали датчиків (акселерометр і гіроскоп) були попередньо оброблені із застосуванням фільтрів шуму. Експериментатори також вели відеозапис, щоб помітити дані вручну. Отриманий набір даних був випадковим чином розділений на дві групи, тестову і тренувальну.

Особливостями спеціалізації Data Science є також наступне:

- повторюваність (курси повторюються кожен місяць) і тривають 4 тижні кожен;
- можливість проходити одночасно кілька курсів;
- можливість заміни деяких курсів на інші в рамках однієї спеціалізації;
- можливість проходити безкоштовно все, крім фінального проекту;
- нахил у бік комерціалізації та отримання верифікованого сертифікату;
- одного разу заплативши за курс, і не закінчивши його, можна пройти його повторно безкоштовно.

Підбиваючи підсумки розгляду особливостей SMOOC відзначимо, що ця нова перспективна форма масових онлайн курсів надає можливість більш глибоко вивчити предметну область. Однак вона передбачає і більш потужну мотивацію студента, якому слід завершити не один курс, а відразу дев'ять! Проблемою залишається і великий відсів учасників.

Так, з 800 тис. студентів, які підписалися на спеціалізацію, завершили принаймні один курс з верифікованим сертифікатом 14 тис. студентів, закінчили всі 9 курсів тільки 266 осіб.

Вивчення особливостей Data Science спеціалізації надає можливість намітити наступні шляхи включення такої форми навчання в традиційний навчальний процес українських ВНЗ: оновлення навчально-методичного забезпечення дисциплін, що читаються, організація самостійної роботи студентів з матеріалами курсів, включення в тематику кваліфікаційних робіт, використання нових засобів і методів, що вивчаються в спеціалізації, в дослідженнях аспірантів і докторантів для аналізу отриманих даних.

Напрями подальших досліджень включають вивчення особливостей інших спеціалізацій масових відкритих онлайн курсів у галузі аналізу даних, зокрема, спеціалізації «Became a Social Scientist» Амстердамського університету, а також практичні кроки для включення такої форми навчання в традиційний навчальний процес ВНЗ України.

Список використаних джерел

1. A review of MOOCs and their assessment tools [Electronic resource]. – Access mode : <http://www.tonybates.ca/2014/11/08/a-review-of-moocs-and-their-assessment-tools/>.
2. Capstone [Electronic resource]. – Access mode : <http://en.wikipedia.org/wiki/Capstone>.
3. Chauhan A. Massive Open Online Courses (MOOCS) : Emerging Trends in Assessment and Accreditation / A. Chauhan // Digital Education Review. – 2014. – No 25. – Access mode : <http://greav.ub.edu/der/index.php/der/article/view/277/491>.
4. Coursera [Electronic resource]. – Access mode : <https://www.coursera.org>.
5. Johns Hopkins University Data Science [Electronic resource]. – Access mode : https://www.coursera.org/specialization/jhudatascience/1?utm_medium=listingPage.
6. Reyes-Ortiz J. Human Activity Recognition Using Smartphones Dataset Version 1.0. [Electronic resource] / Jorge L. Reyes-Ortiz, Davide Anguita, Alessandro Ghio, Luca Oneto. – Access mode: www.smartlab.ws.
7. Heaton J. My thoughts on courses 1-9 of the Johns Hopkins courser data science specialization [Electronic resource] / Jeff Heaton. – Access mode : <http://www.jeffheaton.com/2014/08/my-thoughts-on-courses-1-9-of-the-johns-hopkins-coursera-data-science-specialization/>.
8. What-is-a-capstone-project [Electronic resource]. – Access mode : <https://fanatiq.wordpress.com/2009/07/23/what-is-a-capstone-project/>.
9. Адаменко О. В. Хмарні технології аналізу даних / О. В. Адаменко, Л. Ф. Панченко // Хмарні технології в освіті : матеріали Всеукр. наук.-метод. семінару. – Кривий Ріг : Видавничий відділ КМІ, 2012. – С. 143-144.
10. Calter M. MOOCs and the library: engaging with evolving pedagogy / Calter Mariellen [Electronic resource]. – Access mode : <http://library.ifla.org/160/>.
11. Панченко Л. Ф. Масовий відкритий онлайн курс як альтернативна форма підвищення кваліфікації викладача вищої школи / Л. Ф. Панченко // Education and pedagogical sciences. – 2013. – № 1(156). – С. 19-28.
12. Панченко Л. Ф. Практикум по анализу данных / Л. Ф. Панченко. – Луганск : Изд-во ГУ «ЛНУ имени Тараса Шевченко», 2013. – 269 с.

References (translated and transliterated)

1. A review of MOOCs and their assessment tools [Electronic resource]. – Access mode : <http://www.tonybates.ca/2014/11/08/a-review-of-moocs-and-their-assessment-tools/>.

their-assessment-tools/.

2. Capstone [Electronic resource]. – Access mode : <http://en.wikipedia.org/wiki/Capstone>.

3. Chauhan A. Massive Open Online Courses (MOOCS) : Emerging Trends in Assessment and Accreditation / A. Chauhan // Digital Education Review. – 2014. – No 25. – Access mode : <http://greav.ub.edu/der/index.php/der/article/view/277/491>.

4. Coursera [Electronic resource]. – Access mode : <https://www.coursera.org>.

5. Johns Hopkins University Data Science [Electronic resource]. – Access mode : https://www.coursera.org/specialization/jhudatascience/1?utm_medium=listingPage.

6. Reyes-Ortiz J. Human Activity Recognition Using Smartphones Dataset Version 1.0. [Electronic resource] / Jorge L. Reyes-Ortiz, Davide Anguita, Alessandro Ghio, Luca Oneto. – Access mode: www.smartlab.ws.

7. Heaton J. My thoughts on courses 1-9 of the Johns Hopkins courser data science specialization [Electronic resource] / Jeff Heaton. – Access mode : <http://www.jeffheaton.com/2014/08/my-thoughts-on-courses-1-9-of-the-johns-hopkins-coursera-data-science-specialization/>.

8. What-is-a-capstone-project [Electronic resource]. – Access mode : <https://fanatiq.wordpress.com/2009/07/23/what-is-a-capstone-project/>.

9. Adamenko O. V. Khmarni tekhnolohii analizu danykh [Cloud data analysis technologies] / O. V. Adamenko, L. F. Panchenko // Khmarni tekhnolohii v osviti : materialy Vseukr. nauk.-metod. seminaru. – Kryvyi Rih : Vydavnychiy viddil KMI, 2012. – S. 143-144. (In Ukrainian)

10. Calter M. MOOCs and the library: engaging with evolving pedagogy / Calter Mariellen [Electronic resource]. – Access mode : <http://library.ifla.org/160/>.

11. Panchenko L. F. Masovyi vidkrytyi onlain kurs yak alternatyvna forma pidvyshchennia kvalifikatsii vykladacha vyshchoi shkoly [A massive open online course as an alternative form of continuing education for a high school teacher] / L. F. Panchenko // Education and pedagogical sciences. – 2013. – No. 1(156). – S. 19-28. (In Ukrainian)

12. Panchenko L. F. Praktikum po analizu danykh [The Workshop on Data Analysis] / L. F. Panchenko. – Lugansk : Izd-vo GU “LNU imeni Tarasa Shevchenko”, 2013. – 269 s. (In Russian)